Accuracy of colposcopy-directed punch biopsies: a systematic review and meta-analysis

Underwood M, Arbyn M, Parry-Smith W, De Bellis-Ayres S, Todd R, Redman CW, Moss EL

CRD summary
This review concluded that the high sensitivity and low specificity of colposcopy-directed punch biopsy for high-grade cervical intraepithelial neoplasia might be due to verification bias, as most studies only conducted excision in women with a positive punch biopsy. This conclusion is likely to be reliable.

Authors' objectives
To assess the accuracy of colposcopy-guided punch biopsy for diagnosing high-grade cervical intraepithelial neoplasia.

Searching
MEDLINE, EMBASE and The Cochrane Library were searched for articles from inception to April 2011, and search terms were reported. Additional studies were sought by screening the bibliographies of included articles, reviews and conference abstracts, and by contacting experts.

Study selection
Studies assessing the accuracy of colposcopy-guided punch biopsy in diagnosing high-grade cervical intraepithelial neoplasia (CIN) – CIN2, CIN3 adenocarcinoma in situ, or cervical cancer – were eligible for inclusion. The included studies had to use a reference standard of histological assessment of excision biopsy (large loop excision of the transformation zone, laser or cold knife conisation), or hysterectomy sample, to confirm diagnosis. They had to report sufficient data to populate 2x2 tables of the true-positive, false-positive, false-negative and true-negative biopsy results.

Most of the included studies were retrospective reviews of biopsies taken in colposcopy out-patient clinics. Most of them included women with all grades of cytological abnormality; two included only women with low-grade cytological abnormalities, and one included only women with high-grade cytological abnormalities. A few studies reported that patients had not undergone previous treatment of the cervix. In most studies, women did not receive a biopsy if cervical cancer was obvious at colposcopy.

The authors did not state how many reviewers selected studies for inclusion.

Assessment of study quality
The methodological quality of the included studies was assessed using the QUADAS tool. The authors did not specify how many reviewers assessed quality.

Data extraction
The data were extracted to populate 2x2 contingency tables for the performance of punch biopsy, using a diagnostic threshold of CIN1+ to detect CIN2+ disease, and CIN2+ to detect CIN3+ disease or adenocarcinoma in situ. Sensitivity and specificity estimates, with 95% confidence intervals, were calculated, for each reported combination of test and reference standard threshold.

The data were independently extracted by two reviewers.

Methods of synthesis
A bivariate summary receiver operating characteristic model was used to generate summary estimates of sensitivity and specificity, with 95% confidence intervals, and summary receiver operating characteristic curves, for each combination of test and reference standard threshold. The Cochran Q test was used to assess statistical heterogeneity.

Sensitivity analyses were conducted to assess the effects of biopsy rate (the proportion of patients who received punch biopsy), and interval between test and reference standard (studies where excision was performed during the same surgical procedure as punch biopsy), on measures of test performance.

Results of the review
Thirty-two studies, with 10,598 patients, were included in the review; patients were only included if both biopsy and histology results were available (7,873 women). The results of the quality assessment were not reported, but the authors noted that verification bias was a significant issue, as most studies only conducted excision in women with a positive punch biopsy.

The pooled sensitivity for punch biopsy using a threshold of CIN1+ to diagnose CIN2+ disease was 91.3% (95% CI 85.3 to 94.9) and the pooled specificity was 24.6% (95% CI 16.0 to 35.9), based on data from 25 studies. There was significant between-study heterogeneity for both measures.

The pooled sensitivity for punch biopsy using a threshold of CIN1+ to diagnose CIN3+ disease was 91.1% (95% CI 83.7 to 95.4) and the pooled specificity was 18.2% (95% CI 11.3 to 27.9), based on data from 22 studies. The heterogeneity was not reported.

The pooled sensitivity for punch biopsy using a threshold of CIN2+ to diagnose CIN2+ disease was 80.1% (95% CI 73.2 to 85.6) and the pooled specificity was 63.4% (95% CI 50.9 to 76.7), based on data from 32 studies. There was significant between-study heterogeneity for both measures.

The pooled sensitivity for punch biopsy using a threshold of CIN2+ to diagnose CIN3+ disease was 83.6% (95% CI 74.9 to 89.8) and the pooled specificity was 44.5% (95% CI 34.3 to 55.2), based on data from 27 studies. There was significant between-study heterogeneity for both measures.

For studies where less than 70% of punch biopsies were included in the meta-analyses, and for studies where excision biopsy was performed immediately after punch biopsy (during the same surgical procedure), sensitivity was lower and specificity was higher, than for all studies.

**Authors' conclusions**
The high sensitivity and low specificity of colposcopy-directed punch biopsy for high-grade CIN might be a result of verification bias, as most studies only conducted excision in women with a positive punch biopsy.

**CRD commentary**
The review reported a clear research question and defined some relevant inclusion criteria. Several sources were searched for relevant studies; some sources included unpublished studies and no restrictions were reported. Data extraction included measures to minimise error and bias, but it was not clear whether similar measures were used for study selection and quality assessment. The results of the quality assessment were not reported, but verification bias was discussed, and the authors noted that details for QUADAS criteria were poorly reported.

The authors reported that most studies only conducted excision in women with a positive punch biopsy. Excision biopsy as the reference standard was specified as an inclusion criterion, and it was not clear whether the included studies used any alternative reference standard, such as follow-up, for women with negative punch biopsies, which might have allowed data for these women to be included in the review. Appropriate meta-analytic methods were used and some potential sources of heterogeneity were investigated.

The authors' conclusions are an accurate interpretation of the results reported and are likely to be reliable.

**Implications of the review for practice and research**
The authors did not specify any recommendations for clinical practice and future research.

**Funding**
No funding received.

**Bibliographic details**

**PubMedID**
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.